

# The role of binding entropy in the refinement of protein-ligand docking predictions: analysis based on the use of 11 scoring functions.

Anatoly M. Ruvinsky\*

Center for Bioinformatics, The University of Kansas,  
2030 Becker Drive, Lawrence, KS 66047, USA  
Institute of Spectroscopy, Russian Academy of Sciences,  
Troitsk, Moscow region, 142092 Russia

February 2, 2008

---

\*Tel: 1-(785) 864-1962; Fax: 1-(785) 864-1954; E-mail address: ruvinsky@ku.edu (A.M. Ruvinsky)

## Abstract

We present results of testing of the ability of eleven popular scoring functions to predict native docked positions using a recently developed method [1] for estimation the entropy contributions of relative motions to protein-ligand binding affinity. The method is based on the integration of the configurational integral over clusters obtained from multiple docked positions. We use a test set of 100 PDB protein-ligand complexes and ensembles of 101 docked positions generated by Wang et al [2] for each ligand in the test set. To test the suggested method we compare the averaged root-mean square deviations (RMSD) of the top-scored ligand docked positions, accounting and not accounting for entropy contributions, relative to the experimentally determined positions. We demonstrate that the method increases docking accuracy by 10–21% when used in conjunction with the AutoDock scoring function, by 2–25% with G-Score, by 7–41% with D-Score, by 0–8% with LigScore, by 1–6% with PLP, by 0–12% with LUDI, by 2–8% with F-Score, by 7–29% with ChemScore, by 0–9% with X-Score, by 2–19% with PMF, and by 1–7% with DrugScore. We also compare the performance of the suggested method with the method based on ranking by cluster occupancy only. We analyze how the choice of a RMSD-tolerance and a low bound of dense clusters impacts on docking accuracy of the scoring methods. We derive optimal intervals of the RMSD-tolerance for 11 scoring functions.

**Key words:** protein-ligand docking, binding affinity, entropy, scoring function, cluster occupancy.

## Introduction

The prediction of the experimentally observed positions and conformations of small organic ligands on the surface of macromolecules (e.g. proteins, DNAs) is known as the docking problem. Methods and tools for solving the docking problem attract a great attention in scientific community for many years [1-13]. The accurate and fast solution of the docking problem is of fundamental practical importance for understanding numerous biological process in cells and for the discovery of new drug lead compounds [3-6,11-15]. Docking tests and detailed comparative analysis of the performance of different docking tools [16-23] demonstrate the dependence of docking accuracy on the conformational search methods, the quality of the protein-ligand potentials describing binding enthalpy and scoring methods for estimation of protein-ligand binding entropy.

Scoring functions play an important role in computational studies of protein-ligand structures and of thermodynamics of protein-ligand binding [1,2,6-13], in virtual database screening and drug design [3-7,24-34]. We have recently suggested and validated a novel method to estimate protein-ligand binding entropy [1]. We showed that accounting for the entropy of relative and torsional motions through a configurational integral modifies a commonly used form of scoring functions with a term dependent on occupancy of the clusters obtained from a number of docked positions. The docked positions were generated using AutoDock [24] docking program and then grouped into nonoverlapping clusters in such a way that every cluster contains ligand positions with RMSD less than a pre-set value (a RMSD-tolerance). Ruvinsky and Kozintsev [1] showed that the method essentially improves docking accuracy in comparison with the common method based on ranking by energy when used in conjunction with the AutoDock scoring function. So it is very intriguing and also important to investigate the performance of the method with other scoring functions.

The present article describes results of the application of the method [1] in conjunction with eleven popular scoring functions (AutoDock [24], G-Score [25], D-Score [26], LigScore [27], PLP [28], LUDI [29], F-Score [30], ChemScore [31], X-Score [32], PMF [33], DrugScore [34]) and a test set of 100 PDB protein-ligand complexes developed by Wang et al [2] to predict ligand docked positions. The test set developed by Wang et al [2] essentially differs from a test set of 135 PDB complexes used by Ruvinsky and Kozintsev [1]. The overlap of the test sets consists of three protein-ligand complexes: *2pk4*, *1rbp* and *1rnt*. Wang et al [2] generated ensembles of 101 docked positions for each ligand in the test set (<http://sw16.im.med.umich.edu/software/xtool/>) and scored them by the above mentioned eleven scoring functions. Using these ensembles and the eleven scoring functions modified with the entropy term [1], we reordered docked ligand positions in ensembles. Then we compared the RMSD of top-scored ligand docked positions, accounting and not accounting

for the entropy, relative to the experimentally determined positions.

The organization of this paper is as follows. We derive an expression for the entropy contribution of relative and torsional motions in the Materials and Methods section. Also in the Materials and Methods we describe the test set of protein-ligand PDB complexes and ensembles of docked conformations. In the Results section we compare docking accuracies of calculations with and without entropy contributions in terms of RMSD of the top-scored ligand docked positions relative to the experimentally determined positions. Also in Results we compare the suggested method with the method based on cluster occupancy only. We summarize our conclusions in the final section.

### Materials and Methods: Theory

Protein-ligand binding free energy can be written as [11, 12, 15] (see also [35-47])

$$\Delta G = E_{pl} - E_p - E_l - T \ln \left( \frac{\sigma_l \sigma_p c_o N_a}{\sigma_{pl}} \frac{Z_{pl}}{Z_p Z_l} \right), \quad (1)$$

where  $E_{p,l,pl}$  are the ground energies of protein (p), ligand (l) and protein-ligand complex (pl) in solution;  $N_a$  is the Avogadro number;  $c_o = 1 \text{ mol/l}$ ;  $\sigma_{l,p,pl}$  are the orders of symmetry of ligand, protein and protein-ligand complex (for a nonsymmetrical molecule  $\sigma = 1$ ; if a molecule has 2-fold axis of symmetry  $\sigma = 2$ , etc.);  $Z_{p,l,pl}$  are vibrational partition functions of proteins, ligands and complexes.

Considering only relative protein-ligand motions we can write the protein-ligand binding free energy in the form [1, 11, 12, 15]

$$\Delta G = E_{pl} - E_p - E_l - T \ln \left( \frac{\sigma_l \sigma_p c_o N_a V_B}{\sigma_{pl}} \right), \quad (2)$$

where

$$V_B = \int_{\Gamma} \exp \left( -\frac{U_{pl}(\mathbf{r}, \theta, \varphi, \psi) - E_{pl}}{T} \right) d\mathbf{r} \sin \theta d\theta d\varphi d\psi \quad (3)$$

is the configurational integral of the complex;  $U_{pl}(\mathbf{r}, \theta, \varphi, \psi)$  is the energy of the protein-ligand complex in solution;  $\mathbf{r}$  is the vector of relative translational motions in the complex;  $(\theta, \varphi, \psi)$  are Euler angles of relative orientational motions;  $\Gamma$  is the the region of integration in the 6-dimensional space of  $\mathbf{r}$  and  $(\theta, \varphi, \psi)$ ;  $E_{pl}$  is the minimum of  $U_{pl}(\mathbf{r}, \theta, \varphi, \psi)$  in the region  $\Gamma$ .

Note that to predict the native binding mode corresponding to the minimum of the Exp. (2), we can neglect the contribution of  $E_p + E_l$  to binding free energy. But the absolute value of the binding constant, of course, depends on the energies of the unbound protein and ligand molecules. Thus the binding mode is exactly defined by  $E_{pl}(\Gamma)$  and  $V_B(\Gamma)$ .

This property of the Exp. (3) essentially simplifies docking problem in comparison with the problem of binding affinity prediction and allows searching docked positions using probability distribution functions [48].

Further we follow the method suggested recently [1]. In brief, it depends on the fact that most docking algorithms generate a number of different ligand positions corresponding to different local minima of the protein-ligand energy landscape. To estimate  $V_B$ , we first partition all docked ligand positions (Fig. 1) from a number of runs of an algorithm into non-overlapping clusters in such a way that every cluster contains ligand positions with RMSD less than a definite value ( $0.5 - 4\text{\AA}$ ; see Methods section) relative to the ligand position having minimal energy in the cluster. Now we can consider the clusters as the possible ligand binding modes. Further, we designate the docked ligand position having minimal energy in the cluster as the representative position in the cluster. All ligand positions in the cluster numbered  $i$  can be considered as snapshots of the ligand motion near the representative docked position  $(\mathbf{r}_i, \mathbf{\Omega}_i)$ . The variation intervals of  $(\mathbf{r}, \mathbf{\Omega})$  in the cluster give the estimate of the configurational integral as

$$V_B(\mathbf{r}_i, \mathbf{\Omega}_i) \approx \Gamma_i = [\max(\theta_i) - \min(\theta_i)] [\max(\varphi_i) - \min(\varphi_i)] [\max(\psi_i) - \min(\psi_i)] \\ [\max(x_i) - \min(x_i)] [\max(y_i) - \min(y_i)] [\max(z_i) - \min(z_i)], \quad (4)$$

where  $\max(\mathbf{r}_i, \mathbf{\Omega}_i)$  and  $\min(\mathbf{r}_i, \mathbf{\Omega}_i)$  are the maximum and minimum values of  $(\mathbf{r}, \mathbf{\Omega})$  in the cluster numbered  $i$ .

Omitting the contribution of  $E_p + E_l$  and  $\sigma_l \sigma_p c_o N_a / (8\sigma_{pl} \pi^2)$  we obtain

$$\Delta \tilde{G}_i = E_{pl}(\Gamma_i) - T \ln \Gamma_i, \quad (5)$$

where  $E_{pl}(\Gamma_i)$  is the energy minimum of the protein-ligand complex in the  $i$ -mode. To determine the binding mode we have to determine the set  $\{\Gamma\}$ , calculate the Exp. (5) for all  $\Gamma_i$  and select a representative position having a minimal value of  $\tilde{G}_i$ . Exp. (5) can also be derived by a Monte-Carlo approximation of the configurational integral (3)[1].

The use of  $\Gamma_i \approx N_i v_p$  ( $v_p$  is volume per point in the configurational space,  $N_i$  is the number of conformations in the cluster numbered  $i$ ) converts Exp. (5) into

$$\Delta \tilde{G}_i = E_{pl}(\Gamma_i) - T \ln (N_i v_p) \quad (6)$$

Thus the binding mode is exactly defined by  $E_{pl}(\Gamma_i)$  and  $N_i$ . Further, we use the Exp. (6) to rank the representative positions.

To derive a scoring expression for the method of ranking by cluster occupancy [22, 48, 49] (see also [50-54] and [55, 56] for the using of the method in the studies of the protein folding

and protein-protein docking) we rewrite the Exp. (6) in the following form

$$P(i) = N_i v_p \exp \left( -\frac{E_{pl}(\Gamma_i)}{T} \right) \quad (7)$$

$P(i)$  is proportional to the probability of finding the ligand in the cluster  $i$ . Assuming that  $E_{pl}(\Gamma)$  is a slowly varying function over  $\{\Gamma\}$ , we obtain  $P(i) \sim N_i$ , and the cluster occupancy becomes a single factor identifying the native binding position. Thus the methods of ranking by cluster occupancy or by energy  $E_{pl}$  are the special cases of the more general method based on the Exp. (6).

In the Results section we shall compare results of three scoring methods identifying the binding position as the representative position with the minimal value  $\Delta\tilde{G}$  (Method 1), as the docked position with the minimal energy  $E_{pl}$  (Method 2), and as the representative position in the most occupied cluster (Method 3).

### Materials and Methods: The Test Set

We used the test set of 100 PDB protein-ligand complexes [2]: 1bbz, 4xia, 8xia, 2xim, 1fkf, 1fkb, 1hvr, 1tet, 2cgr, 1abf, 1apb, 7abp, 5abp, 8abp, 9abp, 1abe, 1bap, 6abp, 1e96, 1add, 2ak3, 1adb, 9aat, 1bzm, 1cbx, 2ctc, 3cpa, 1cla, 3cla, 4cla, 2csc, 5cna, 1af2, 1dr1, 1dhf, 1drf, 1ela, 7est, 3fx2, 2gbp, 1hsl, 2qwd, 2qwe, 2qwf, 2qwg, 2qwc, 2qwb, 1mnc, 1exw, 1apw, 1apt, 1bxo, 1fmo, 2pk4, 1inc, 4sga, 5sga, 5p21, 1rbp, 1rgk, 6rnt, 1rgl, 1rnt, 1zzz, 1yyy, 1b5g, 1ba8, 1bb0, 2sns, 1sre, 7tln, 4tln, 1tmn, 2tmn, 3tmn, 5tln, 1tlp, 1etr, 1ets, 1d3d, 1d3p, 1a46, 1a5g, 1bcu, 1tha, 4tim, 6tim, 7tim, 1bra, 1tnj, 1pph, 1tnk, 1tnh, 1tni, 1ppc, 1tng, 3ptb, 1tnl, 1bhf, 2xis. All these entries have resolution better than  $2.5\text{\AA}$ . Wang et al [2] generated an ensemble of 101 docked conformations for each ligand in the test set. One of the conformations corresponds to the experimentally observed native conformation of the ligand. RMSD distributions in the conformational ensembles spread from  $0\text{\AA}$  to  $15\text{\AA}$ . For RMSD-tolerance of  $2\text{\AA}$  ensembles consist of 30 – 70 distinctive conformational clusters.

To analyze an ability of the scoring functions to predict the experimentally observed conformations eleven scoring functions were applied to score the conformational ensembles [2]. We used the scored ensembles of each ligand and applied the Exp. (6) to test the ability of the suggested method to predict the native conformation. We varied the RMSD-tolerance from  $0.5$  to  $4\text{\AA}$ .

### Results

For correct use of the Exp. (6) it is necessary to keep in mind that it is based on the estimate of the configurational integral. Thus only clusters with high occupancy should be

scored with Exp. (6). To differentiate between dense and sparse clusters, we introduce a low bound  $N_{lb}$  of dense clusters. Only clusters with  $N_i \geq N_{lb}$  are scored with Exp. (6). If all clusters have occupancy lower than  $N_{lb}$ , we select the most occupied cluster as the cluster of the binding position, but if several clusters have the same occupancy lower than  $N_{lb}$ , we compare them using Exp. (6). Further we consider the percentage of the top-ranked solutions within a RMSD of  $2\text{\AA}$  of the experimental result and designate this value as the success rate (SR). Success rates of three scoring methods used in conjunction with 11 scoring functions are given in Fig. 2-12 for the low bound of dense clusters equal to 3, 4, 5 and 10.

### **Force field based scoring functions**

#### **AutoDock**

As the RMSD-tolerance increases from  $1\text{\AA}$  to  $4\text{\AA}$ , Method 1 based on Exp. (6) applied with the AutoDock scoring function improves the success rate by 3 – 13% relative to the results of ranking by energy (Method 2) not accounting for the entropy effect (Fig. 2-a). SR of the Method 1 reaches maximum of 75% for the RMSD-tolerance of  $2.5\text{\AA}$  and the low bound of 4. SR of ranking by cluster occupancy (Method 3) reaches maximum of 73% for the RMSD-tolerance of  $2.5\text{\AA}$ . For the RMSD-tolerance of 0.5, 1, 3.5 and  $4\text{\AA}$  ranking by cluster occupancy is less successful than ranking by energy.

#### **G-Score**

Fig. 2-b shows that Method 1 applied together with the G-Score scoring function improves the success rate by 2 – 25% relative to the results of "bare" G-Score (Method 2). SR of the Method 1 reaches maximum of 67% for the RMSD-tolerance of  $2.5\text{\AA}$  and the low bound of 5 and 10. SR of ranking by cluster occupancy reaches maximum of 70% for the RMSD-tolerance of  $2.5\text{\AA}$  also. For all values of the RMSD-tolerance energy ranking is a worse predictor than ranking based on Exp. (6) and by cluster occupancy.

#### **D-Score**

Applying Exp. (6) in conjunction with the D-Score scoring function improves essentially the success rate by 7 – 41% relative to the results of the Method 2 (Fig. 2-c). SR of the Method 1 reaches maximum of 67% for the RMSD-tolerance of  $2.5\text{\AA}$  and the low bound of 5 and 10. SR of the Method 3 has a maximum of 69% for the RMSD-tolerance of  $2.5\text{\AA}$ . The worst results of scoring ( $SR = 26\%$ ) for all values of the RMSD-tolerance are obtained for Method 2.

### **Empirical scoring functions**

#### **LigScore**

We found that as the RMSD-tolerance increases from  $2\text{\AA}$  to  $4\text{\AA}$ , Method 1 applied with the LigScore scoring function improves the success rate by  $0 - 8\%$  relative to the results of Method 2 (Fig. 2-d). For the RMSD-tolerance  $< 2\text{\AA}$ , the difference between Method 2 using "bare" LigScore and Methods 1 and 2 is  $1 - 8\%$  and  $6 - 13\%$ . SR of Method 1 using Exp. (6) reaches maximum of  $82\%$  for the RMSD-tolerance of  $2.5\text{\AA}$  and the low bound of 3 and 4. SR of Method 3 using cluster occupancy reaches maximum of  $74\%$  for the RMSD-tolerance of  $2\text{\AA}$  and  $2.5\text{\AA}$ . For the RMSD-tolerance  $> 2.5\text{\AA}$ , Method 2 with energy ranking outperforms Method 3 by  $1 - 8\%$ , but worse than Method 1 by  $0 - 8\%$ .

### PLP

Docking accuracy of Method 1 applied with PLP scoring function (Fig. 2-e) is comparable with the accuracy of Method 2 using energy ranking for the case of the RMSD-tolerance lower than  $2\text{\AA}$ , but becomes better by  $1 - 6\%$  for the case of the RMSD-tolerance higher than  $2\text{\AA}$ . SR of the Method 1 reaches maximum of  $82\%$  for the RMSD-tolerance of  $3.0\text{\AA}$  and the low bound of 4. SR of Method 3 using ranking by cluster occupancy reaches maximum of  $79\%$  for the RMSD-tolerance of  $3.0\text{\AA}$ . For two values of the RMSD-tolerance of  $3\text{\AA}$  and  $3.5\text{\AA}$  ranking by cluster occupancy slightly outperforms energy ranking by  $3\%$  and  $1\%$ , but loses  $0 - 3\%$  to the results of ranking by Method 1. For other values of the RMSD-tolerance Methods 1, 2 show better docking results than Method 3.

### LUDI

Fig. 2-f shows that as the RMSD-tolerance increases from  $0.5\text{\AA}$  to  $4\text{\AA}$ , Method 1 applied with the LUDI scoring function improves the success rate by  $0 - 12\%$  relative to the results of Method 2. SR of Method 1 using Exp. (6) reaches maximum of  $79\%$  for the RMSD-tolerance of  $3.0\text{\AA}$  and the low bound of 10. SR of Method 3 reaches maximum of  $77\%$  for the RMSD-tolerance of  $3.0\text{\AA}$ . Method 2 outperforms by  $2 - 7\%$  Method 3 only in two cases of the RMSD-tolerance of  $0.5, 1.0\text{\AA}$ .

### F-Score

The results of ranking docked positions on the basis of Methods 1, 2 and 3 in conjunction with F-Score scoring function are shown in Fig. 2-g. We can see that as the RMSD-tolerance increases from  $1.5\text{\AA}$  to  $4\text{\AA}$ , Method 1 improves the success rate by  $2 - 8\%$  relative to the results of Method 2. SR of Method 1 reaches maximum of  $82\%$  for the RMSD-tolerance of  $3.0\text{\AA}$  and the low bound of 3, 5, 10. SR of Method 3 using ranking by cluster occupancy reaches maximum of  $77\%$  for the RMSD-tolerance of  $3.0\text{\AA}$  and  $3.5\text{\AA}$ . For the RMSD-tolerance lower than  $2\text{\AA}$  Method 2 ( $SR = 74\%$ ) outperforms Method 3 by  $3 - 9\%$ .



## ChemScore

Fig. 2-h shows that there is a significant improvement of docking accuracy for Methods 1 and 3 applied with ChemScore scoring function, in comparison with the results of the Method 2, choosing the energy top-ranked position to identify the binding position. Methods 1 and 3 outperform the Method 2 by 6 – 29% and 8 – 32% accordingly. SR of Method 1 using Exp. (6) reaches maximum of 64% for the RMSD-tolerance of 1.5, 2.0 Å and the low bound of 4, 5. SR of Method 3 reaches maximum of 67% for the RMSD-tolerance of 2.0 Å.

## X-Score

Fig. 2-i shows that Methods 1 and 3 applied with the X-Score scoring function improve the success rate by 0 – 9% for the RMSD-tolerance from 1.0 Å to 4 Å, and 0 – 8% for the RMSD-tolerance from 1.5 Å to 4 Å relative to the results of Method 2 not accounting for the entropy effect. SR of Method 1 reaches maximum of 75% for the RMSD-tolerance of 2.5 Å and the low bound of 4, 5, 10. SR of Method 3 using ranking by cluster occupancy reaches maximum of 74% for the RMSD-tolerance of 2.5 Å. For the RMSD-tolerance of 0.5, 1 and 4 Å energy ranking ( $SR = 66\%$ ) slightly outperforms by 1 – 3% results of Method 3.

## Knowledge-based scoring functions

### PMF

Fig. 2-j illustrates results of applying ranking Methods 1, 2 and 3 in conjunction with the PMF scoring function. We observe that Methods 1 and 3 outperform by 2 – 19% and 2 – 20% the results of the common Method 2 using energy ranking ( $SR = 52\%$ ). SR of Method 1 using Exp. (6) reaches maximum of 71% for the RMSD-tolerance of 2.5 Å and the low bound of 5. SR of Method 3 using ranking by cluster occupancy reaches maximum of 72% for the RMSD-tolerance of 2.5 Å.

## DrugScore

As the RMSD-tolerance increases from 1.0 Å to 4 Å, Method 1 applied with the DrugScore scoring function improves the success rate by 1 – 7% (Fig. 2-k) relative to the results of Method 2 ( $SR = 72\%$ ). SR of using Method 1 reaches maximum of 79% for the RMSD-tolerance of 3.0 Å and the low bound of 4, and the RMSD-tolerance of 4.0 Å and the low bound of 10. SR of Method 3 reach maximum of 75% for the RMSD-tolerance of 2 Å and 3 Å. For the RMSD-tolerance lower than 1.5 Å and equal to 4 Å energy ranking (Method 2) outperforms by 2 – 7% results of Method 3 using ranking by cluster occupancy.

## Discussions

The success rate of ranking using Exp. (6) (Method 1) shows a bell-shape curve behavior (Fig. 2) for all scoring functions except LigScore (Fig. 2-d). It means that 10 of 11 scoring functions closely describe protein-ligand energy landscapes in the test set. Tails of the bell-shape functions approach a value of the success rate of energy ranking neglecting the entropy effect. This is the result of reduction of the suggested method to the method of ranking by energy for a very small or large RMSD-tolerance of cluster size. Indeed, for a very low RMSD-tolerance all clusters contain a single docked position and ranking using the Method 1 and ranking by energy (Method 2) are became identical. In the limit of a very large RMSD-tolerance, only one cluster exists and thus ranking by Method 1 and Method 2 give the same results for the success rate. The bell-shape curve behavior for PMF is not so evident (Fig. 2-j) as for AutoDock (Fig. 2-a), D-Score (Fig. 2-c), LUDI (Fig. 2-f) or ChemScore (Fig. 2-h). However, it can be detected averaging SR over different values of the low bound of the cluster size for every value of the RMSD-tolerance or by following rhombus and stars on Fig. 2-j. The SR behavior for Method 1 applied in conjunction with LigScore as a function of the RMSD-tolerance and the low bound of the cluster size has the same character for a very low and large values of the RMSD-tolerance, but differs in the most interesting range of intermediate values of the RMSD-tolerance. It has two clear extrema - a minimum for the RMSD-tolerance of 1Å and a maximum for the RMSD-tolerance of 2.5Å.

It is interesting to note that all scoring functions demonstrate even behavior of a maximum as a function of the RMSD-tolerance. Thus applying Method 1 we can vary the RMSD-tolerance in pre-set intervals keeping a level of docking accuracy averaged over the low bound of the cluster size. So AutoDock allows one to vary the RMSD-tolerance from 1.5Å to 3Å (Fig. 2-a), G-Score - from 1.5Å to 2.5Å (Fig. 2-b), D-Score - from 1.5Å to 2.5Å (Fig. 2-c), LigScore - from 2.5Å to 4Å (Fig. 2-d), PLP - from 2.5Å to 4Å (Fig. 2-e), LUDI - from 2Å to 3Å (Fig. 2-f), F-Score - from 2.5Å to 4Å (Fig. 2-g), ChemScore - from 1.5Å to 2.5Å (Fig. 2-h), X-Score - from 2Å to 3Å (Fig. 2-i), PMF - from 1.5Å to 3.5Å (Fig. 2-j), DrugScore - from 2.5Å to 4Å (Fig. 2-k). Considering optimal values of the RMSD-tolerance and the low bound of the cluster size, we found that Method 1 outperforms Method 2 in docking accuracy by 10–21% when used in conjunction with the AutoDock scoring function, by 2 – 25% with G-Score, by 7 – 41% with D-Score, by 0 – 8% with LigScore, by 1 – 6% with PLP, by 0 – 12% with LUDI, by 2 – 8% with F-Score, by 7 – 29% with ChemScore, by 0 – 9% with X-Score, by 2 – 19% with PMF, and by 1 – 7% with DrugScore. These results are the unambiguous evidence of improving docking accuracy by accounting for the entropy of relative and torsional motions.

The success rate of scoring over cluster occupancy (Method 3) also shows the bell-shape

curve behavior for AutoDock (Fig. 2-a), G-Score (Fig. 2-b), D-Score (Fig. 2-c), ChemScore (Fig. 2-h), X-Score (Fig. 2-i) and PMF (Fig. 2-j). For PLP (Fig. 2-e), LigScore (Fig. 2-d), LUDI (Fig. 2-f), F-Score (Fig. 2-g), DrugScore (Fig. 2-k) curves of SR as a function of the RMSD-tolerance differ from the bulb function behavior and show several extrema. Considering optimal values of the RMSD-tolerance, we found that Method 3 outperforms Method 2 in docking accuracy by 11% when used in conjunction with the AutoDock scoring function, by 28% with G-Score, by 43% with D-Score, by 3% with PLP, by 10% with LUDI, by 3% with F-Score, by 32% with ChemScore, by 8% with X-Score, by 20% with PMF, and by 3% with DrugScore. Best results of Method 3 applied with LigScore coincide with the results of ranking by LigScore energy (Method 2).

It is interesting to note that SRs of ranking using Method 3 and Method 1 show good correlation for AutoDock, D-Score, G-Score, LUDI, ChemScore and X-Score. It means that using these scoring functions and applying Methods 1 or 3 we choose the same top-scored representative position satisfying simultaneously to the following inequalities

$$N_1 > N_i \text{ and } N_1 \exp\left(-\frac{E_{pl}(\Gamma_1)}{T}\right) > N_i \exp\left(-\frac{E_{pl}(\Gamma_i)}{T}\right), \quad (8)$$

where 1 is the cluster number of the top-scored representative position. Other representative positions are numbered  $i \neq 1$ . Using Ineq. (8) we obtain

$$E_{pl}(\Gamma_1) - E_{pl}(\Gamma_i) < T \ln \frac{N_1}{N_i} \quad (9)$$

On condition  $E_{pl}(\Gamma_1) > E_{pl}(\Gamma_i)$ , top-scored representative positions reside not in deepest energy wells. Using Ineq. (9) we can estimate the maximal difference between depths of energy wells in protein-ligand energy landscape as  $T \ln(N_1/\min(N_i)) = 2.8\text{kcal/mol}$  for  $T = 300\text{K}$ ,  $N_1 = 99$  and  $\min(N_i) = 1$ . If  $E_{pl}(\Gamma_1) < E_{pl}(\Gamma_i)$ , then top-scored representative positions reside in the deepest and mostly occupied energy wells.

## Conclusions

We presented results of testing 11 popular scoring functions on 100 protein-ligand complexes using the recently suggested method [1], accounting for binding entropy of relative motions in a protein-ligand complex, and two other commonly used methods of ranking by energy or cluster occupancy. We rigorously showed that both methods of ranking by cluster occupancy and by energy are the special cases of the more general method accounting for binding entropy. We applied the three ranking methods to the conformational ensembles generated by Wang et al [2] and compared efficiencies of the methods in terms of the percentage of the top-ranked solutions within a RMSD of  $2\text{\AA}$  of the experimental result (the success rate).

We demonstrated that the method based on Exp. (6) predicts native position significantly better than the top energy ranking, when used in conjunction with D-Score, ChemScore, AutoDock, G-Score, or PMF, and moderately or slightly better when used with LigScore, PLP, LUDI, F-Score, X-Score, or DrugScore. The presented results prove that the method can be applied together with all types of current force fields, empirical scoring functions or knowledge-based potentials.

For the most of tested scoring functions we observed strong correlations between docking accuracies of two methods of ranking using Exp. (6) and ranking by cluster occupancy. These correlations suggest that for these potentials the near-native conformations, in comparison with far-native ones, have the greatest number of neighboring conformations within a RMSD-tolerance. Similar trends were observed recently in studies of protein-ligand docking [22, 48, 49], predictions of protein-protein complexes [55, 56] and studies of protein folding landscape [50, 51, 52, 53]. Also this concept was used by Xiang et al [54] for loop prediction. The authors suggested to rank conformations by a standard energy term together with a RMSD-dependent term that favors conformations that have many neighbors in configurational space. We believe that the method to treat the entropy effect using Exp. (6) should give statistical-thermodynamic explanations of these results and prove useful for future studies of protein folding and protein-protein docking.

### Acknowledgements

I thank to I.A. Vakser for careful review of the manuscript.

## References

- [1] Ruvinisky, A.M.; Kozincev, A.V. J Comp Chem 2005, 26, 1089-1095.
- [2] Wang, R.; Lu, Y.; Wang, S. J Med Chem 2003, 46, 2287-2303.
- [3] Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. Nat Rev Drug Discov 2004, 3, 935-949.
- [4] Stockwell, B.R. Nature 2004, 432, 846-854.
- [5] Shoichet, B.K. Nature 2004, 432, 862-865.
- [6] Abagyan, R.; Totrov, M. Curr Op Chem Biol 2001, 5, 375-382.
- [7] Brooijmans, N.; Kuntz, I. Annu Rev Biophys Biomol Struct 2003, 32, 335-373.

- [8] Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Prot Str Func Gen* 2002, 47, 409-443.
- [9] Gohlke, H.; Klebe, G. *Angew Chem Int Ed* 2002, 41, 2644-2676.
- [10] Cole, J.C.; Murray, C.W.; Nissink, J.W.M.; Taylor, R.D.; Taylor, R. *Prot Str Func Bio* 2005, 60, 325-332.
- [11] Ajay, Murcko, M.A. *J Med Chem* 1995, 38, 4953-4967.
- [12] Gilson, M.K.; Given, J.A.; Head, M.S. *Chem Biol* 1997, 4, 87-92.
- [13] Brooks, C.L. III; Karplus, M.; Pettitt, B.M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics. Advances in Chemical Physics* 71; John Wiley and Sons: New York, 1988.
- [14] Mackerell, A.D. *J Comp Chem* 2004, 25, 1584-1604.
- [15] Gilson, M.K.; Given, J.A.; Bush, B.L.; McCammon, J.A. *Biophys J* 1997, 72, 1047-1069.
- [16] Bissantz, C.; Folkers, G.; Rognan, D. *J Med Chem* 2000, 43, 4759-4767.
- [17] Stahl, M.; Rarey, M. *J Med Chem* 2001, 44, 1035-1042.
- [18] Perez, C.; Ortiz, A.R. *J Med Chem* 2001, 44, 3768-3785.
- [19] Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. *Prot Str Funct Bio* 2004, 57, 225-242.
- [20] Kontoyianni, M.; McClellan, L.M.; Sokol, G.S. *J Med Chem* 2004, 47, 558-565.
- [21] Wang, W.; Donini, O.; Reyes, C.M.; Kollman, P.A. *Annu Rev Biophys Biomol Struct* 2001; 30: 211-243.
- [22] Källblad, P.; Mancera, R.L.; Todorov, N.P. *J Med Chem* 2004, 47, 3334-3337.
- [23] Ruvinsky, A.M.; Kozincev, A.V. *Biophys Chem* 2005, 115, 255-261.
- [24] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. *J Comput Chem* 1998, 19, 1639-1662.
- [25] Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. *J Mol Biol* 1997, 267, 727-748.
- [26] Ewing, T.J.A.; Makino, S.; Skillman, A.G.; Kuntz, I.D. *J Comput-Aided Mol Des* 2001, 15, 411-428.

- [27] Cerius2, version 4.6; Accelrys Inc.; <http://www.accelrys.com/>
- [28] Gehlhaar, D.K.; Verkhivker, G.M.; Rejto, P.A.; Sherman, C.J.; Fogel, D.B.; Fogel, L.J.; Freer, S.T. *Chem Biol* 1995, 2, 317-324.
- [29] Böhm, H.J. *J Comput- Aided Mol Des* 1994, 8, 243-256.
- [30] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J Mol Biol* 1996, 261, 470-489.
- [31] Eldridge, M.D.; Murray, C.W.; Auton, T.R.; Paolini, G.V.; Mee, R.P. *J Comput-Aided Mol Des* 1997; 11: 425-445.
- [32] Wang, R.; Lai, L.; Wang, S. *J Comput-Aided Mol Des* 2002, 16, 11-26.
- [33] Muegge, I.; Martin, Y.C. *J Med Chem* 1999, 42, 791-804.
- [34] Gohlke, H.; Hendlich, M.; Klebe, G. *J Mol Biol* 2000, 295, 337-356.
- [35] Gö, N.; Gö, M.; Scheraga, H.A. *Proc Natl Acad Sci USA* 1968, 59, 1030-1037.
- [36] Karplus, M.; Kushick, J. *Macromol* 1981, 14, 325-332.
- [37] Hermans, J.; Shankar, S. *Isr J Chem* 1986, 27, 225-227.
- [38] Finkelstein, A. V.; Janin, J. *Prot Eng* 1989, 3, 1-3.
- [39] Roux, B.; Nina, M.; Pomes, R.; Smith, J.C. *Biophys J* 1996; 71: 670-681.
- [40] Amzel, L. M. *Protein Struct Funct Genet* 1997, 28, 144-148.
- [41] Simonson T.; Archontis, G.; Karplus, M. *Accts Chem Res* 2002, 35, 430-437.
- [42] Gohlke, H.; Case, D.A. *J Comp Chem* 2004, 25, 238 - 250.
- [43] Luo, H.; Sharp, K. *Proc Natl Acad Sci USA* 2002, 99, 10399-10404.
- [44] Swanson, J.M.J.; Henchman, R.H.; McCammon, J.A. *Biophys J* 2004, 86, 67-74.
- [45] Chen, W.; Chang C.-E.; Gilson, M.K. *Biophys J* 2004, 87, 3035-3043.
- [46] Hsu, S.T.; Peter, C.; van Gunsteren, W.F.; Bonvin, A.M. *Biophys J* 2005, 88, 15-24
- [47] Carlsson, J.; Åqvist, J. *J Phys Chem B* 2005, 109, 6448-6456.
- [48] Ruvinsky, A.M.; Kozincev, A.V. *Prot Str Func Bio* 2006, 62, 202-208.

- [49] Rosenfeld, R.J.; Goodsell, D.S.; Musah, R.A.; Morris, G.M.; Goodin, D.B.; Olson, A.J. J Comp-Aid Mol Des 2003, 17, 525-536.
- [50] Shortle, D.; Simons, K.T.; Baker, D. Proc Natl Acad Sci USA 1998, 95, 11158-11162.
- [51] Bonneau, R.; Tsai, J.; Ruczinski, I.; Chivian, D.; Rohl, C.; Strauss, C.E.M.; Baker, D. Prot Str Funct Gen Suppl 2001; 5: 119-126.
- [52] Wang, K.; Fain, B.; Levitt, M.; Samudrala, R. BMC Str Biol 2004, 4, 1-18.
- [53] Zhang, Y.; Skolnick, J. J Comput Chem 2004, 25, 865-871.
- [54] Xiang, Z.; Soto, C.S.; Honig, B. Proc Natl Acad Sci USA 2002; 99: 7432-7437.
- [55] Tovchigrechko, A.; Vakser, I.A. Prot Sci 2001, 10, 1572-1583.
- [56] Comeau, S.R.; Gatchell, D.W.; Vajda, S.; Camacho, C.J. Bioinf 2004, 20, 45-50.

## Legend to figures.

1. The clustering scheme. Small circles are local minima of the protein-ligand energy landscape found in docking.

2. Percentage of the top-ranked representative solutions within a RMSD of  $2\text{\AA}$  from the experimentally determined position, scored using Exp. (6) and a) AutoDock, b) G-Score, c) D-Score, d) LigScore, e) PLP, f) LUDI, g) F-Score, h) ChemScore, i) X-Score, j) PMF, k) DrugScore, for the low bound of dense clusters equal to 3 (circles), 4 (triangles), 5 (rhombus) and 10 (stars) as a function of the RMSD-tolerance. Solid line corresponds to the success rate of the scoring function neglecting the entropy effect. Rectangles connected with by dash line correspond to the success rate of ranking by cluster occupancy.



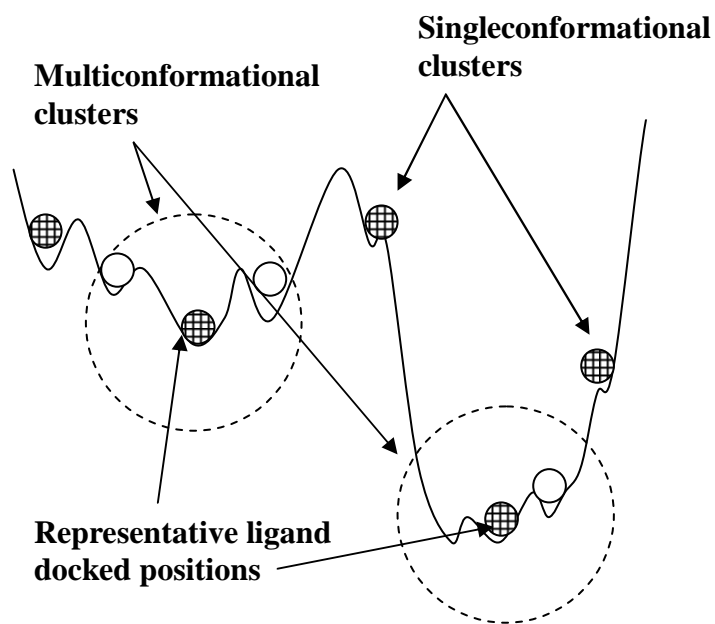


Fig. 1.

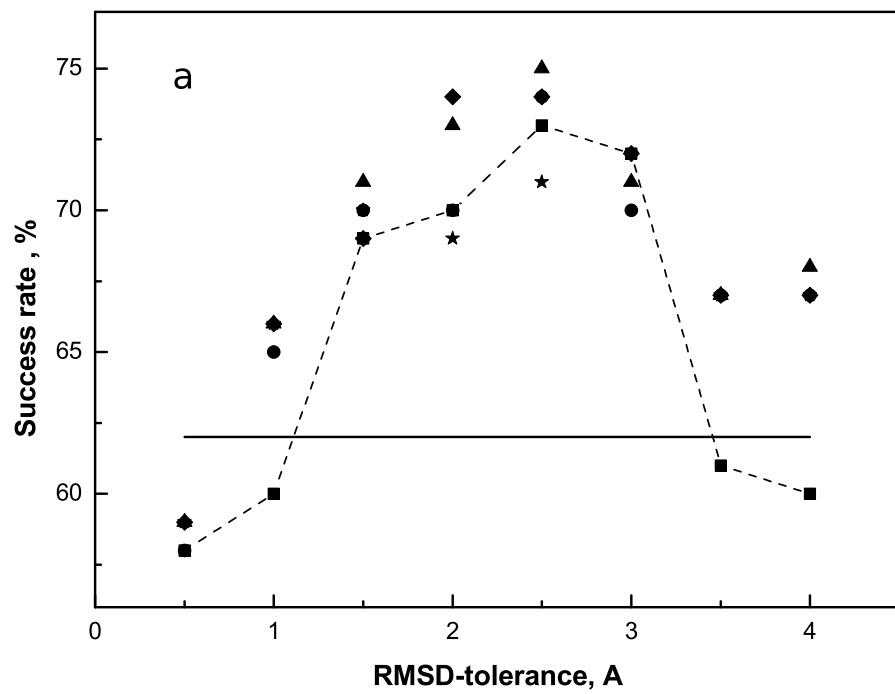


Fig. 2 a

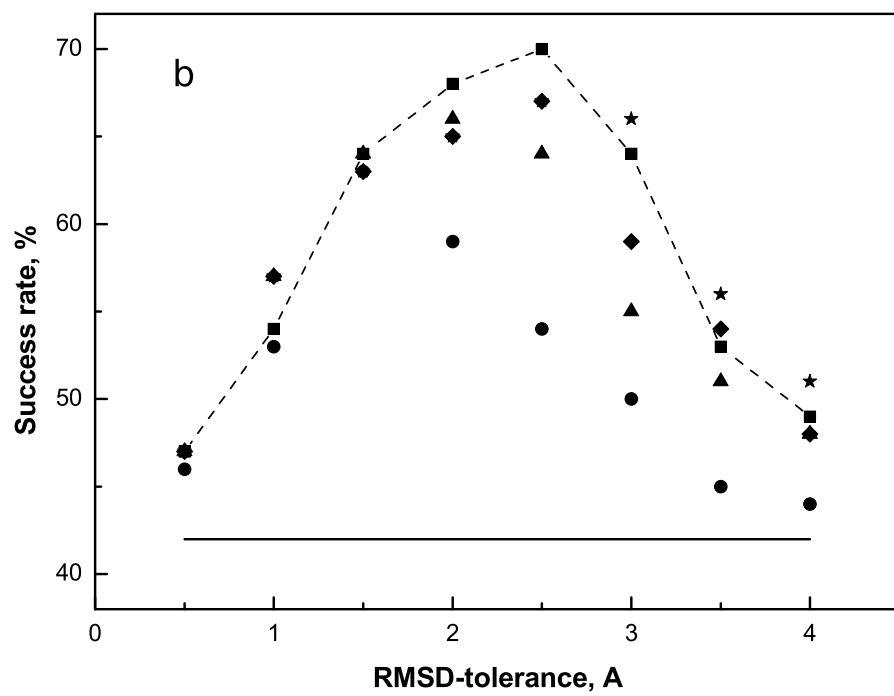


Fig. 2 b

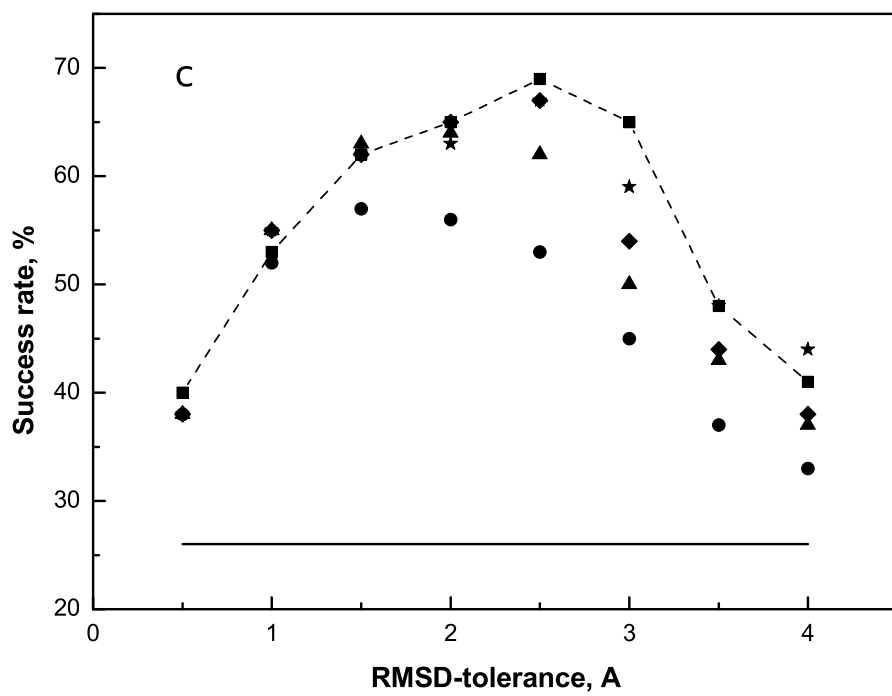


Fig. 2 c

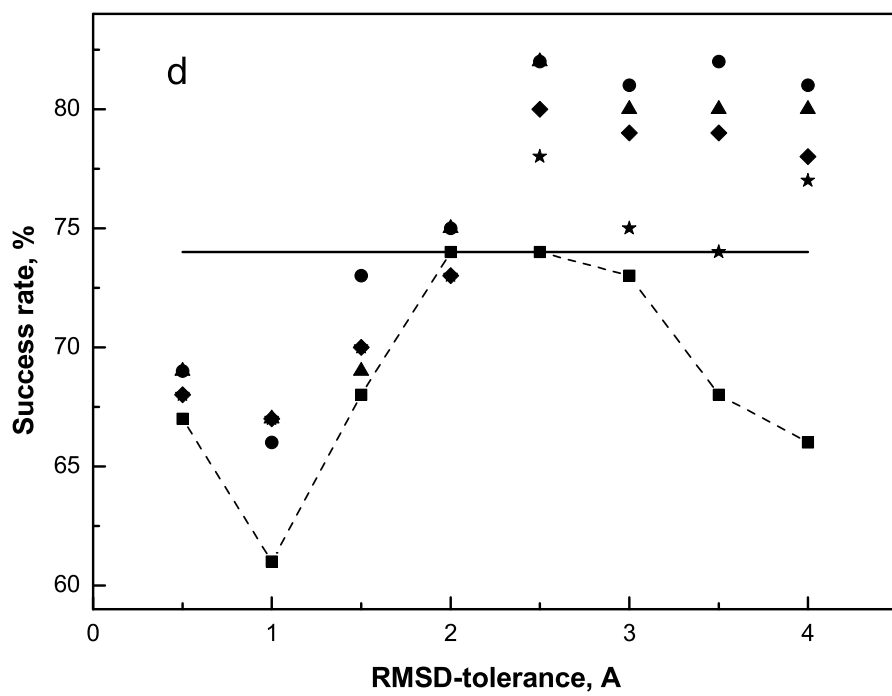


Fig. 2 d

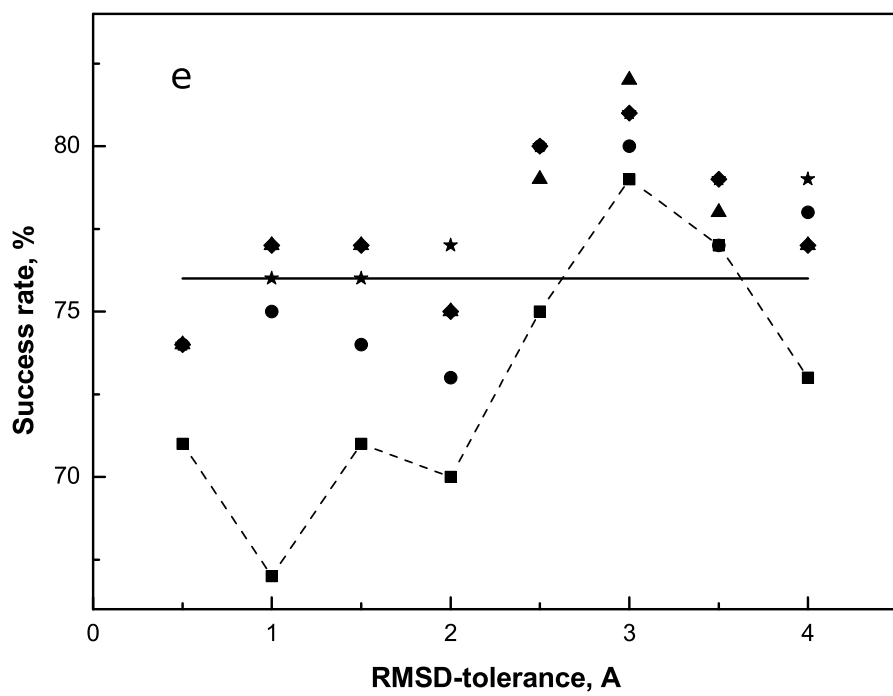


Fig. 2 e

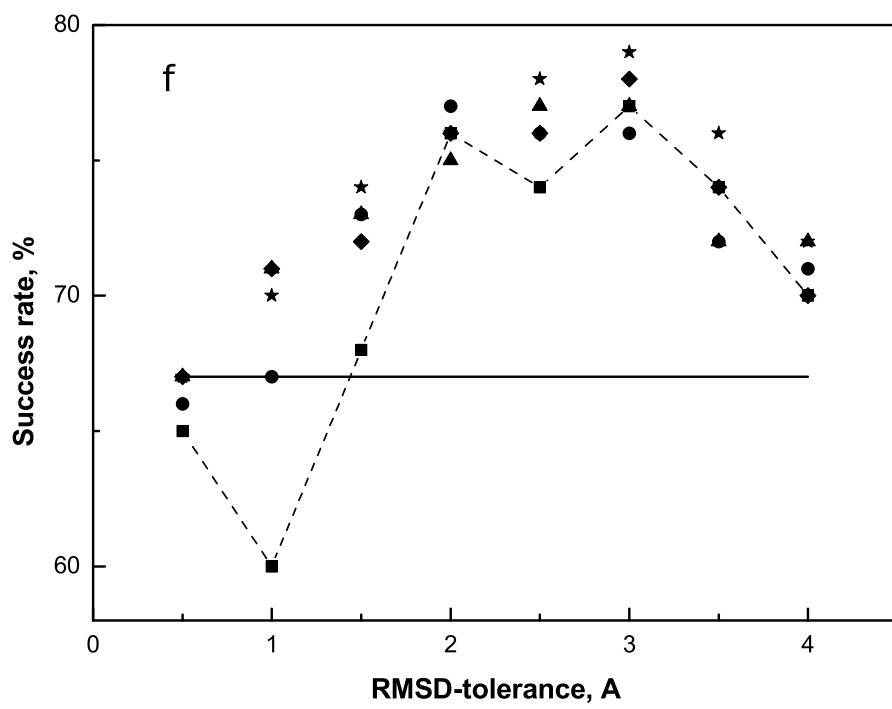


Fig. 2 f

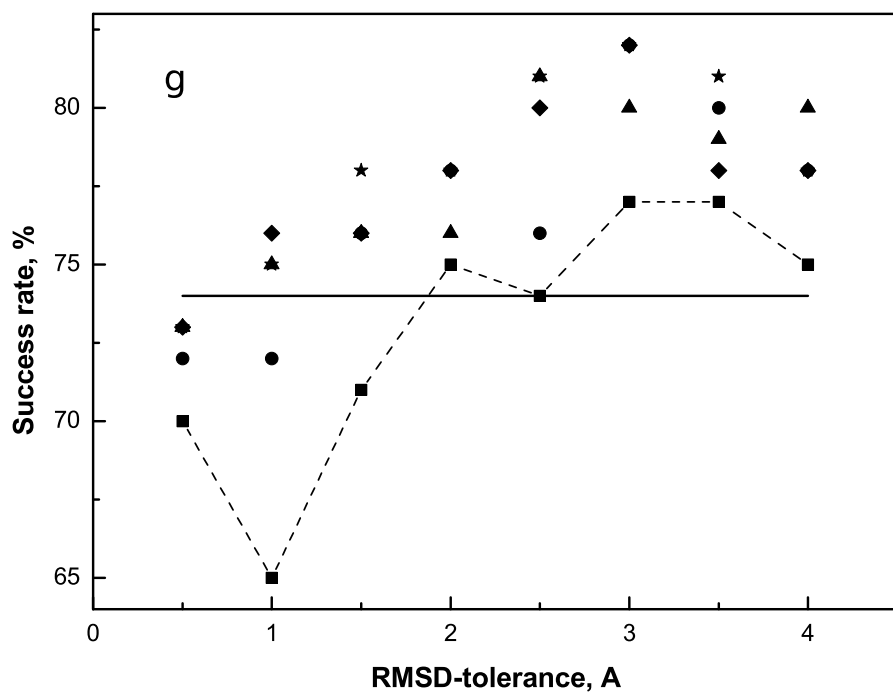


Fig. 2 g



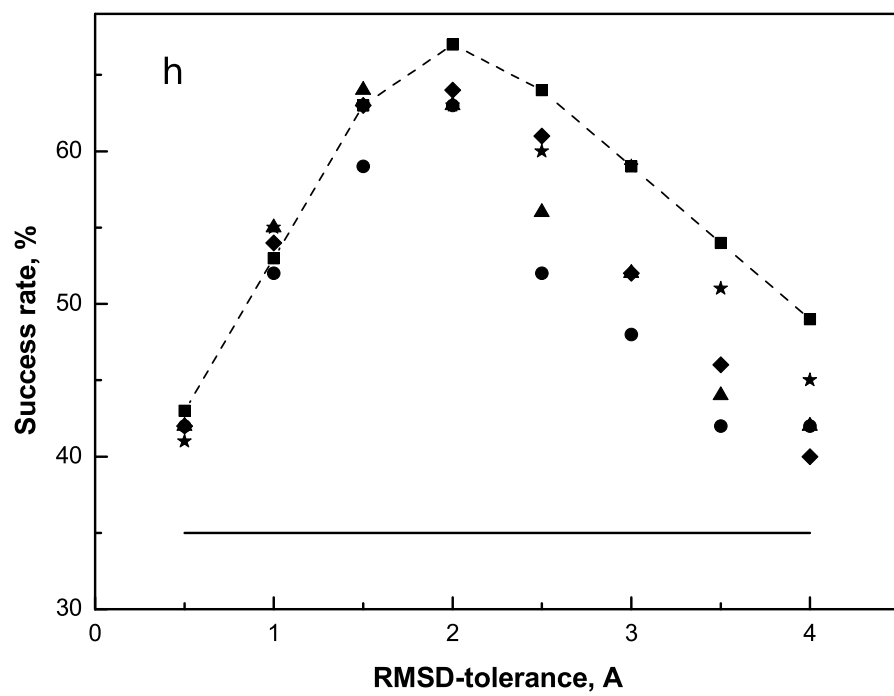


Fig. 2 h

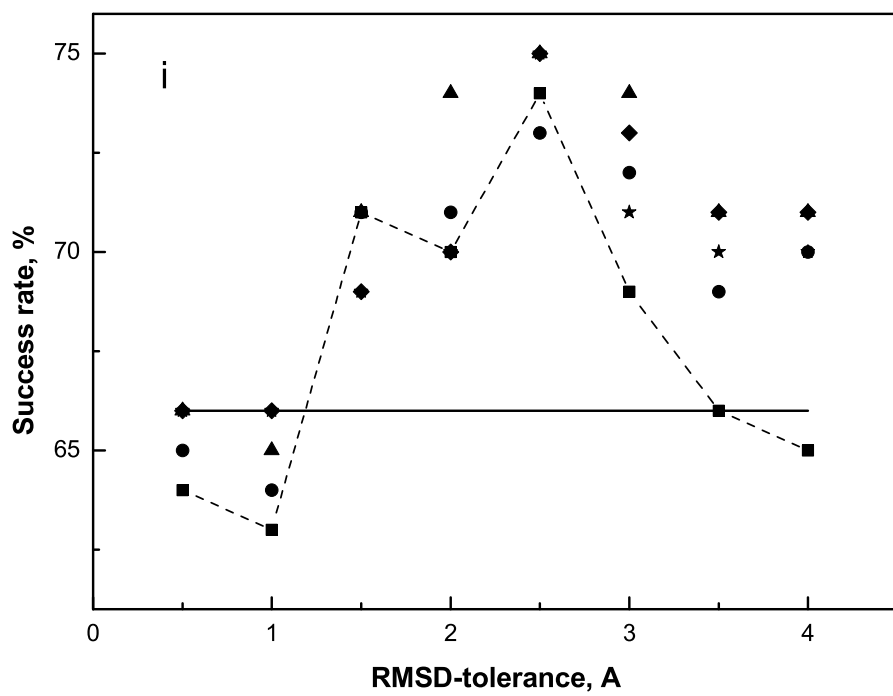


Fig. 2 i

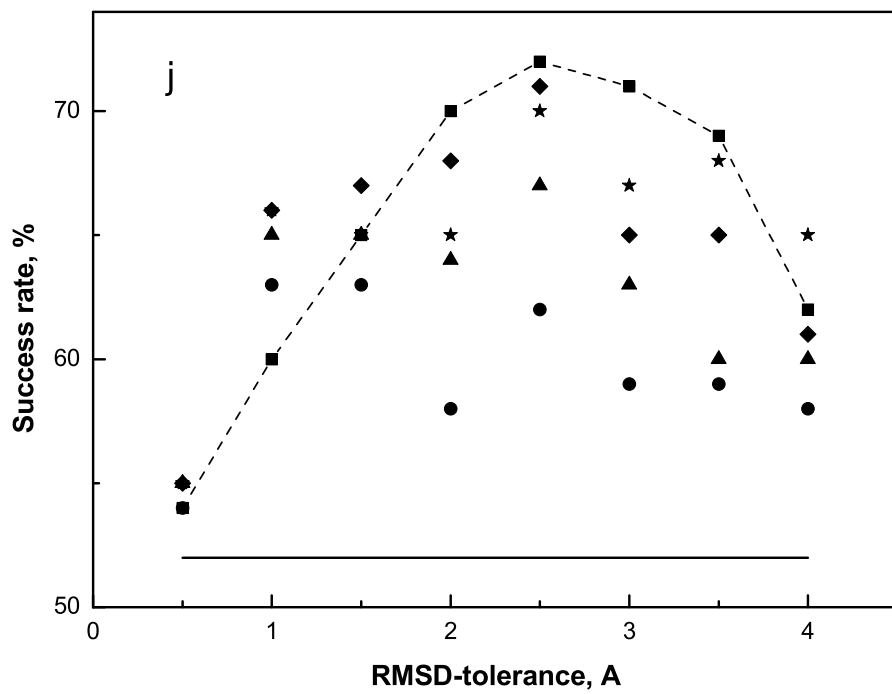


Fig. 2 j

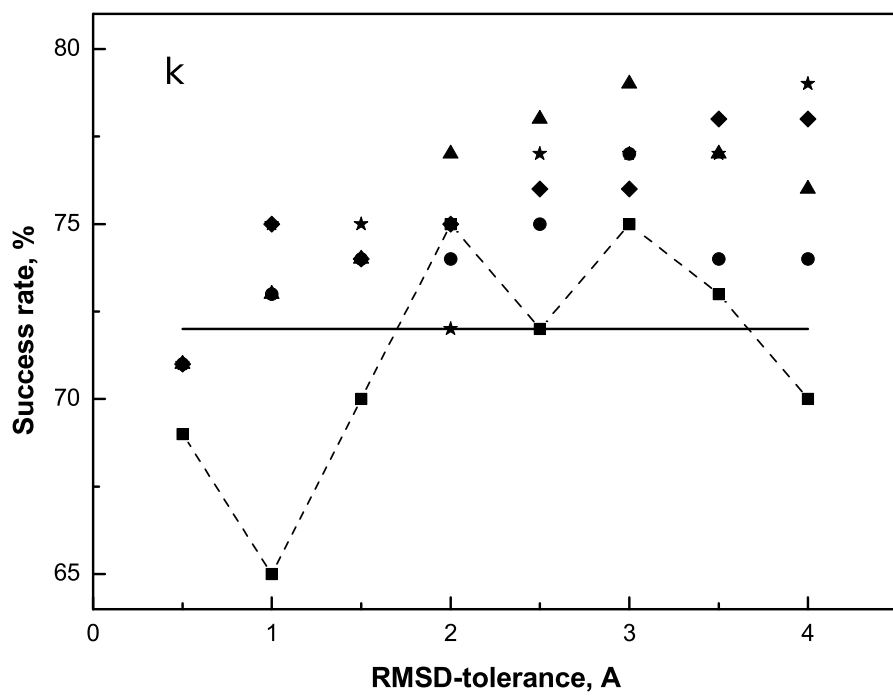


Fig. 2 k